

CLAIMS

5 What is claimed is:

SUB A17

- 10 1. A method of developing a domain-specific analytic application having at least one predefined data mining model, the method comprising the steps of:
- identifying a business problem to be solved;
- selecting a data mining algorithm appropriate for solving the business problem;
- 15 defining data schema for use as inputs and outputs to and from the mining algorithm, the data schema including input data schema and output data schema; and
- 20 defining a data mining model dependent upon the data schema, defining a data mining model resulting in the creation of a predefined data mining model;
- whereby a domain-specific analytic application is
- 25 developed, the analytic application having at least one predefined data mining model.
2. The method of claim 1 wherein the analytic application comprises the capabilities of:
- 30 production training the predefined data mining model using the historical data in the input data schema, wherein use of the capability of production training the

predefined data mining model results in creation of a
production trained data mining model; and

production scoring production data by use of the
5 production trained data mining model.

3. The method of claim 2 wherein the capability of
production training the predefined data mining model
further comprises the capability of operating the
10 predefined data mining model in training mode using end
user historical data in the input data schema.

4. The method of claim 2 wherein the capability of
production scoring production data by use of the
15 production trained data mining model further comprises
the capability of applying the production trained data
mining model to historical data stored in input schema.

5. The method of claim 2 wherein the capability of
20 production scoring production data by use of the
production trained data mining model further comprises
the capability of applying the production trained data
mining model to production data stored read from an end
user's production database.

25 6. The method of claim 1 wherein the analytic application
further comprises the capability of populating the input
data schema with historical data.

30 7. The method of claim 6 wherein the capability of
populating the input data schema with historical data
further comprises the capabilities of extracting from
historical data values of prediction data fields and

writing the values of the prediction data fields into the input data schema for the data mining model.

8. The method of claim 1 wherein identifying a business
5 problem to be solved further comprises identifying a
a usiness problem capable of expression through the use
of referents that are defined in a specific computational
domain.
- 10 9. The method of claim 1 wherein selecting a mining
algorithm appropriate for solving the business problem
further comprises selecting a radial basis function
algorithm for value prediction.
- 15 10. The method of claim 1 wherein selecting a mining
algorithm appropriate for solving the business problem
further comprises selecting a neural value prediction
algorithm.
- 20 11. The method of claim 1 wherein selecting a mining
algorithm appropriate for solving the business problem
further comprises selecting a demographic clustering
algorithm.
- 25 12. The method of claim 1 wherein selecting a mining
algorithm appropriate for solving the business problem
further comprises selecting a neural clustering
algorithm.
- 30 13. The method of claim 1 wherein selecting a mining
algorithm appropriate for solving the business problem
further comprises selecting a tree classification
algorithm.

14. The method of claim 1 wherein selecting a mining
algorithm appropriate for solving the business problem
further comprises selecting a neural classification
5 algorithm.

15. The method of claim 1 wherein selecting a mining
algorithm appropriate for solving the business problem
further comprises selecting an associations algorithm.
10

16. The method of claim 1 wherein defining data schema
for the mining algorithm further comprises the steps of:

15 selecting from historical data for inclusion in
input data schema predictor fields capable of supporting
the use of a data mining algorithm in predicting the
value of a prediction field; and

20 selecting for inclusion in output data schema at
least one prediction field.

17. The method of claim 1 wherein defining data schema
for the mining algorithm further comprises selecting
25 for inclusion in output schema sufficient key fields to
comprise a unique key for identification in production
data of storage locations for the output data from the
data mining algorithm.

30 18. The method of claim 1 wherein defining a data mining
model based on the data schema further comprises
establishing in a data structure comprising the data
mining model definition values for fields defining the

model.

19. The method of claim 18 wherein fields defining the model comprise:

5 a field representing the number of consecutive records to select from the input data schema to be used for training;

a field representing the number of consecutive
10 records to select from the input data schema to be used for development scoring;

a field limiting the number of times the data mining model goes through its input data in training mode;

15 a field limiting the number of fitting centers created by the mining data mining algorithm at each pass through the input data;

20 a field indicating the minimum number of records to be assigned to a region;

a field identifying at least one predictor field;
And

25 a field identifying a prediction field.

20. The method of claim 1 wherein defining a data mining model based on the data schema further comprises the
30 steps of:

establishing in a data structure comprising the data mining model definition values for fields defining the

09826662-1

model; and

development scoring historical data wherein test
output data is created; and

5

testing the test output data for accuracy wherein an
accuracy valuation is created;

wherein the steps of establishing definition values,
10 development scoring, and testing are repeated until the
accuracy valuation meets a predetermined accuracy
requirement.

21. A system for developing a domain-specific analytic
15 application having at least one predefined data mining
model, the system comprising:

means for identifying a business problem to be
Solved;

20

means for selecting a data mining algorithm
appropriate for solving the business problem;

means for defining data schema for use as inputs and
25 outputs to and from the mining algorithm, the data schema
including input data schema and output data schema; and

means for defining a data mining model dependent
upon the data schema, wherein use of the means for
30 defining a data mining model results in creation of a
predefined data mining model;

wherein use of the said means for identifying a

business problem, means for selecting a data mining algorithm, means for defining data schema, and means for defining a data mining model results in development of a domain-specific analytic application, the analytic
5 application having at least one predefined data mining model.

22. The system of claim 21 wherein the analytic application comprises the capabilities of:

10 production training the predefined data mining model using the historical data in the input data schema, wherein use of the capability of production training the predefined data mining model results in creation of a production trained data mining model; and

15 production scoring production data by use of the production trained data mining model.

23. The system of claim 22 wherein the capability of
20 production training the data mining model further comprises the capability of operating the data mining model in training mode using end user historical data in the input data schema.

25 24. The system of claim 22 wherein the capability of production scoring production data by use of the production trained data mining model further comprises the capability of applying the production trained data mining model to historical data stored in input schema.

30 25. The system of claim 22 wherein the capability of production scoring production data by use of the production trained data mining model further comprises

the capability of applying the production trained datamining model to production data stored read from an end user's production database.

5 26. The system of claim 21 wherein the analytic application further comprises the capability of populating the input data schema with historical data.

27. The system of claim 26 wherein the capability of
10 populating the input data schema with historical data further comprises the capabilities of extracting from historical data values of prediction data fields and writing the values of the prediction data fields into the input data schema for the data mining model.

15 28. The system of claim 21 wherein means for identifying a business problem to be solved further comprises means for identifying a business problem capable of expression through the use of referents that are defined in a
20 specific computational domain.

29. The system of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business problem further comprises means for selecting a radial
25 basis function algorithm for value prediction.

30. The system of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business problem further comprises means for selecting a neural
30 value prediction algorithm.

31. The system of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business

problem further comprises means for selecting a demographic clustering algorithm.

32. The system of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business problem further comprises means for selecting a neural clustering algorithm.

10 33. The system of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business problem further comprises means for selecting a tree classification algorithm.

15 34. The system of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business problem further comprises means for selecting a neural classification algorithm.

20 35. The system of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business problem further comprises means for selecting an associations algorithm.

25 36. The system of claim 21 wherein means for defining data schema for the mining algorithm further comprises:

means for selecting from historical data for inclusion in input data schema predictor fields capable of supporting the use of a data mining algorithm in predicting the value of a prediction field; and

means for selecting for inclusion in output data

schema at least one prediction field.

37. The system of claim 21 wherein means for defining data schema for the mining algorithm further comprises means for selecting for inclusion in output schema sufficient key fields to comprise a unique key for identification in production data of storage locations for the output data from the data mining algorithm.

38. The system of claim 21 wherein means for defining a data mining model based on the data schema further comprises means for establishing in a data structure comprising the data mining model definition values for fields defining the model.

39. The system of claim 38 wherein fields defining the model comprise:

a field representing the number of consecutive records to select from the input data schema to be used for training;

a field representing the number of consecutive records to select from the input data schema to be used for development scoring;

a field limiting the number of times the data mining model goes through its input data in training mode;

a field limiting the number of fitting centers created by the mining data mining algorithm at each pass through the input data;

a field indicating the minimum number of records to

be assigned to a region;

a field identifying at least one predictor field;

and

5 a field identifying a prediction field.

40. The system of claim 21 wherein means for defining a data mining model based on the data schema further comprises:

10

means for establishing in a data structure comprising the data mining model definition values for fields defining the model; and

15

means for development scoring historical data wherein test output data is created; and

means for testing the test output data for accuracy wherein an accuracy valuation is created;

20

wherein the means for establishing definition values, means for development scoring, and means for testing are capable of repeated use until the accuracy valuation meets a predetermined accuracy requirement.

25

41. A computer program product for developing a domain-specific analytic application having at least one predefined data mining model, the computer program product comprising:

30

a recording medium;

means, recorded on the recording medium, for identifying a business problem to be solved;

means, recorded on the recording medium, for selecting a data mining algorithm appropriate for solving the business problem;

5

means, recorded on the recording medium, for defining data schema for use as inputs and outputs to and from the mining algorithm, the data schema including input data schema and output data schema; and

10

means, recorded on the recording medium, for defining a data mining model dependent upon the data schema, wherein use of the means for defining a data mining model results in creation of a predefined data mining model;

15

wherein use of the said means for identifying a business problem, means for selecting a data mining algorithm, means for defining data schema, and means for defining a data mining model results in development of a domain-specific analytic application, the analytic application having at least one predefined data mining model.

20

42. The computer program product of claim 21 wherein the analytic application comprises the capabilities of:

25

production training the predefined data mining model using the historical data in the input data schema, wherein use of the capability of production training the predefined data mining model results in creation of a production trained data mining model; and

30

production scoring production data by use of the
production trained data mining model.

43. The computer program product of claim 22 wherein the
5 capability of production training the data mining model
further comprises the capability of operating the data
mining model in training mode using end user historical
data in the input data schema.

10 44. The computer program product of claim 22 wherein the
capability of production scoring production data by use
of the production trained data mining model further
comprises the capability of applying the production
trained data mining model to historical data stored in
15 input schema.

45. The computer program product of claim 22 wherein the
capability of production scoring production data by use
of the production trained data mining model further
20 comprises the capability of applying the production
trained data mining model to production data stored read
from an end user's production database.

46. The computer program product of claim 21 wherein the
25 analytic application further comprises the capability of
populating the input data schema with historical data.

47. The computer program product of claim 26 wherein the
capability of populating the input data schema with
30 historical data further comprises the capabilities of
extracting from historical data values of prediction data
fields and writing the values of the prediction data
fields into the input data schema for the data mining

model.

48. The computer program product of claim 21 wherein means for identifying a business problem to be solved
5 further comprises means, recorded on the recording medium, for identifying a business problem capable of expression through the use of referents that are defined in a specific computational domain.

10 49. The computer program product of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business problem further comprises means, recorded on the recording medium, for selecting a radial basis function algorithm for value prediction.

15 50. The computer program product of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business problem further comprises means, recorded on the recording medium, for selecting a neural
20 value prediction algorithm.

51. The computer program product of claim 21 wherein means for selecting a mining algorithm appropriate for solving the business problem further comprises means,
25 recorded on the recording medium, for selecting a demographic clustering algorithm.

52. The computer program product of claim 21 wherein means for selecting a mining algorithm appropriate for
30 solving the business problem further comprises means, recorded on the recording medium, for selecting a neural clustering algorithm.

53. The computer program product of claim 21 wherein
means for selecting a mining algorithm appropriate for
solving the business problem further comprises means,
recorded on the recording medium, for selecting a tree
5 classification algorithm.

54. The computer program product of claim 21 wherein
means for selecting a mining algorithm appropriate for
solving the business problem further comprises means,
10 recorded on the recording medium, for selecting a neural
classification algorithm.

55. The computer program product of claim 21 wherein
means for selecting a mining algorithm appropriate for
15 solving the business problem further comprises means,
recorded on the recording medium, for selecting an
associations algorithm.

56. The computer program product of claim 21 wherein
20 means for defining data schema for the mining algorithm
further comprises:

means, recorded on the recording medium, for
selecting from historical data for inclusion in input
25 data schema predictor fields capable of supporting the
use of a data mining algorithm in predicting the value of
a prediction field; and

means, recorded on the recording medium, for
30 selecting for inclusion in output data schema at least
one prediction field.

57. The computer program product of claim 21 wherein

means for defining data schema for the mining algorithm further comprises means, recorded on the recording medium, for selecting for inclusion in output schema sufficient key fields to comprise a unique key for
5 identification in production data of storage locations for the output data from the data mining algorithm.

58. The computer program product of claim 21 wherein means for defining a data mining model based on the data
10 schema further comprises means, recorded on the recording medium, for establishing in a data structure comprising the data mining model definition values for fields defining the model.

15 59. The computer program product of claim 38 wherein fields defining the model comprise:

a field representing the number of consecutive records to select from the input data schema to be used
20 for training;

a field representing the number of consecutive records to select from the input data schema to be used for development scoring;
25

a field limiting the number of times the data mining model goes through its input data in training mode;

a field limiting the number of fitting centers
30 created by the mining data mining algorithm at each pass through the input data;

a field indicating the minimum number of records to

be assigned to a region;

a field identifying at least one predictor field;

and

5 a field identifying a prediction field.

60. The computer program product of claim 2 wherein means for defining a data mining model based on the data schema further comprises:

10

means, recorded on the recording medium, for establishing in a data structure comprising the data mining model definition values for fields defining the model; and

15

means, recorded on the recording medium, for development scoring historical data wherein test output data is created; and

20 means, recorded on the recording medium, for testing the test output data for accuracy wherein an accuracy valuation is created;

wherein the means for establishing definition
25 values, means for development scoring, and means for testing are capable of repeated use until the accuracy valuation meets a predetermined accuracy requirement.